DOCUMENT RESUME

ED 322 971 JC 900 468

AUTHOR Preece, Mary
TITLE The Reliability, Validity, and Usefulness of Student Evaluations for the Purpose of Improving Teaching in Postsecondary Institutions: Annotated Bibliography.
PUB DATE Jan 90
NOTE 21p.; Graduate seminar paper, University of Toronto.
PUB TYPE Reference Materials - Bibliographies (131) -- Information Analyses (070)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Annotated Bibliographies; Cognitive Development; Community Colleges; Course Evaluation; Educational Objectives; Literature Reviews; Response Style (Tests); Student Behavior; *Student Evaluation of Teacher Performance; *Teacher Effectiveness; Teacher Influence; *Test Reliability; *Test Validity; Two Year Colleges

ABSTRACT

This literature review and annotated bibliography on student evaluations of teacher performance begins by reviewing the issues involved and the results of relevant research. The introductory section indicates that student evaluations of teaching effectiveness are commonly solicited in Ontario postsecondary institutions as diagnostic feedback to faculty about their teaching performance. Their use is regularly criticized by teachers, who question the validity and reliability of the data collected. Results of studies regarding the reliability and validity of students' evaluations are highly contradictory. Part of the problem lies in the preconceived biases of some who study student ratings, while variations in the evaluation instruments can make effective assessment difficult. Similarly, if there is a discrepancy between the grade expected by a student, and the actual grade received, this can significantly effect an evaluation's validity. Research shows that instructor sex, rank, and grading standards have minimal impact on student ratings, while there exists a positive correlation between student examination results and student instructor ratings. Student ratings have been found to be highly valid indices of achievement of attitudinal and motivational goals of education, and reasonably valid indices of achievement of cognitive goals. In general, research on student evaluations suggests that students are competent to report on teaching effectiveness and that the data is reliable and, when taken in context, valid. The annotated bibliography lists 30 papers, reports, book chapters, journal articles, and monographs, published between 1971 and 1988. (JMC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED322971

The Reliability, Validity, and Usefulness of Student Evaluations for the Purpose of Improving Teaching in Postsecondary Institutions

Annotated Bibliography

Curriculum Guidelines and Student Evaluation
1363

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

M. Preece

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☒ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Prepared for: Prof. Merlin Wahlstrom
Prepared by: Mary Preece

January, 1990

BEST COPY AVAILABLE

JC 900468

The Reliability, Validity, and Usefulness of Student Evaluations for the Purpose of Improving Teaching in Postsecondary Institutions

Introduction

Students' evaluations of teaching effectiveness are commonly solicited in Ontario postsecondary institutions. The purpose of these evaluations may be to provide: (a) a measure of teaching effectiveness to be used in a summative capacity to support hire/fire decisions, (b) diagnostic feedback to faculty about the effectiveness of their teaching, (c) information for students to use in the selection of courses and instructors, and (d) an outcome on a process description for research on teaching (Marsh, 1984). It is the use of students' evaluations as diagnostic feedback to faculty about the effectiveness of their teaching (formative evaluation) which is the prime concern of this report.

The use of students' evaluations to measure teaching effectiveness is often criticized by teachers. Typical concerns, as described by Aleamoni (1974), include the suggestion that (1) students cannot make consistent judgements about teaching, (2) most student rating schemes are nothing more than a popularity contest, (3) students are not able to make accurate judgements until they have been away from the course, (4) student rating forms are unreliable and invalid, (5) extraneous variables or conditions such

as class size, schedule, terms, student sex, major, and level can affect student ratings, (6) the grade a student receives in the course is highly correlated with his rating of the course and instructor, and (7) the use of student ratings is questionable for the improvement of instruction.

Marsh (1984) suggests that there are several reasons for the concerns identified by Aleamoni. First, college faculty have little or no formal training in teaching. They find themselves in a position where their job security is dependent on their classroom skills. Any procedure used to evaluate teaching effectiveness is, therefore, threatening and highly criticized. Also, there are no clearly defined criteria of effective teaching, particularly when there continues to be considerable debate about the validity of student ratings. Much of the debate about student evaluations is based on ill-founded fears regarding the validity and reliability of the data collected.

Study results regarding the reliability and validity of students' evaluations of teaching vary from "reliable, valid, and useful" to "unreliable, invalid and useless". Marsh (1984) believes that part of the problem lies in the preconceived biases of those who study student ratings. A second part of the problem lies in unrealistic expectations of what student evaluations can and should be able to do. Another part of the problem lies in the plethora of ad hoc instruments based upon varied item content and untested psychometric properties. Finally, part of the

problem lies in the fragmentary approach to the design of both student evaluation instruments and the research based upon them.

The objective of this report is to produce an annotated bibliography which will provide an overview of the literature concerned with the validity, reliability and usefulness of students' evaluations of effective teaching for the formative evaluation of postsecondary faculty.

Reliability

Reliability refers to the consistency of evaluation results. If quite similar scores are obtained when the same evaluation instrument is administered to the same group on two different occasions, it can be concluded that the results have a high degree of reliability from one occasion to another. With evaluation results, consistency is an important quality, but only if it is accompanied by evidence of validity. Reliability is a necessary but not a sufficient condition for validity (Gronlund, 1985).

Student ratings are reliable and consistent (Cross et al., 1988; Seldin, 1980). The reliability coefficients are lower for different courses or even for different sections of the same course (Cross et al., 1988; Murray, 1972). Murray (1972) determined that student ratings are consistent across question items, raters, time periods, and successive years for the same course. "Typical measures of split-half reliability consistently show that the instruments employed in student ratings have a high reliability.

They have been checked and rechecked in different studies" (Seldin, 1980, p. 40).

Validity

Validity refers to the extent to which the results of an evaluation procedure serve the particular uses for which they were intended (Gronlund, 1985). The specific use to be made of the results and the soundness of the proposed interpretations are validity concerns. It is possible for a test to be reliable without being valid, but not for a test to be valid without being reliable.

There are different types of validity. Content validity is important to describe how an individual performs on a representative sample of the domain of tasks under consideration. Establishing content validity involves (1) clearly defining the domain of tasks to be measured, and (2) constructing, or selecting, a representative set of test tasks (Gronlund, 1985). A Table of Specifications may be used to check on content validity.

Construct Validity concerns the interpretation of evaluation results in terms of some psychological trait or quality. A construct is a psychological quality that is assumed to exist in order to explain some aspect of behaviour (Gronlund, 1985). Examples of constructs are intelligence, scientific attitude, critical thinking, reading comprehension, study skills, and mathematical aptitude. The process of determining construct

validity involves: (1) identifying the constructs presumed to account for effective teaching, (2) deriving hypotheses regarding effective teaching from the theory underlying each construct, and (3) verifying the hypothesis by logical and empirical means.

A number of factors tend to influence the validity of evaluation interpretations. Some of these influences can be found in the evaluation instrument itself, in the administration and scoring of the evaluation, in the atypical responses of students, and in the nature of the group to whom the evaluation is being administered.

Questions that must be considered when addressing the issue of validity include: (1) Are the ratings biased? (2) Do the ratings agree with ratings of other relevant groups, and (3) Do students rate highest those instructors from whom they learn the most? (Aubrecht, 1984; Murray, 1980; and Cross et al., 1988)

Are the ratings biased? In statistics a bias is a latent influence that disturbs an analysis. In students' evaluations of teaching effectiveness, a bias is a systematic error in student ratings.

Must research has been conducted to determine what attributes or conditions may have an adverse effect on student ratings. Among the bias concerns are student characteristics such as age, sex, level of study, and personality; student expectations; class size, subject matter, and whether the course is a major or an elective; and instructor sex, rank, and grading standards (Seldin, 1984).

The age, sex, level of study, and the personality of the student have been found to have little impact on students' evaluations of effective teaching (Seldin, 1984).

Student expectation has been determined to be the most significant variable. Students' grades influence their evaluations if there is a discrepancy between the grade expected and the grade actually received (Bausel et al., 1972; Holmes, 1972; and Kennedy, 1975). Kennedy (1975) suggested in his study that there is also a small correlation between students' grades and students' evaluations of instruction. In his study, students with A and B grades evaluated their instructors more favourably than those with C and D grades. This finding, however, was not confirmed in the other research articles. Student expectations are also related to faculty reputation. Perry et al. (1979) determined that a positive instructor reputation results in a higher rating by students. It is possible to compensate for any bias resulting from disconfirmed grade expectancies by communicating clearly with students about their progress and the criteria used for student evaluation (Holmes, 1972). Bias introduced into evaluations as a result of faculty reputation is harder to counteract.

The issue of class size, subject matter, and whether the course is a major or an elective is yet another potential source of bias. Centra (1972) noted that the subject matter being taught can have a slight effect on ratings. The above factors, however, generally have little influence on student ratings (Seldin, 1984).

Holmes, D.S. (1972). "Effects of Grades and Disconfirmed Grade Expectancies on Students' Evaluations of Faculty Performance." Journal of Educational Psychology, 63, 130-133.

Half of the students in this study who deserved and expected A's or B's were given their expected grades, while half were given a grade one step lower than expected. After receiving the grades, the students filled out the Teaching Assessment Blank. A 2 x 2 analysis of variance revealed that the evaluation of the instructor was lowered on only 1 of the 19 items as a function of differences in grades, but evaluation on 10 of the 19 items were lowered as a function of the unexpected lowering of grades. If students' grades disconfirm their expectations, the students tend to **deprecate** the instructor's teaching performance in areas other than his grading system. Differences in grades did not influence students' evaluations of their instructor when these differences were expected. It was the author's opinion that the effect was due to the students' attempts to justify their unexpectedly low grades. If students are kept adequately informed of their proficiency in a course and if they receive grades accurately reflecting their proficiency, the possibility of disconfirmed expectancies will be decreased, and any bias due to disconfirmed expectancies will be decreased also. In general, it appears that students' ratings may be an effective means of assessing students' opinions of their instructors.

Jacobs, Lucy Cheser (1987). "University Faculty and Students' Opinions of Student Ratings." Indiana Studies in Higher Education, No. 55. Bloomington: Indiana University

This study surveyed faculty and students of Indiana University to ascertain their attitudes toward the evaluation of faculty by means of student ratings. Results indicated that (a) faculty generally support the use of student ratings, (b) faculty believe the ratings have helped them improve their teaching, (c) students say that they take faculty ratings seriously, and (d) among factors considered important by students were how much they learned in the course and the instructor's fairness, impartiality, organization, preparation, rapport with students, and expertise in the field. It was concluded that student ratings are used appropriately and that they represent an efficient and valid source of data on teaching.

Kennedy, W. Robert (1975). "Grades Expected and Grades Received - Their Relationship to Students' Evaluations of Faculty Performance." Journal of Educational Psychology, 67, 109-115.

The relationship between students' ratings of faculty performance and the grades they expected and those they received were discussed in this report. The study results suggest that student appraisal of faculty tends to be invalidated by the influence of grades - that students' reactions to grades, whether positive or negative, bias their perceptions of the teacher. Students who received either A or B grades were significantly higher in their evaluations of the teacher than those who received C and D grades. This suggests a correlation between student grades and ratings students assigned to faculty. Since faculty may use the ratings as feedback to alter their teaching habits, an adequate instrument with a basic validity must be the goal of the instrument developer. Validity may be positively affected by the use of a two-stage approach to teacher/course evaluation instrument construction. In the first stage, it would be determined which teacher behaviours or course aspects are unique to specific content areas. In the second part, appropriate items would be developed to appraise the behaviours in each area. This practice would be a first step toward establishing content validity for rating instruments to enhance their value as a feedback mechanism.

Kierstead, Diane; D'Agostino, Patti; and Dill, Heidi (1988). "Sex Role Stereotyping of College Professors: Bias in Students' Ratings of Instructors." Journal of Educational Psychology. 880, 342-344.

The impact of (a) social contact between instructor and students, (b) instructor's facial expression, and (c) instructor's sex on students' ratings of instruction was investigated. Hypothetical scenarios were presented in which these factors varied while behaviours directly related to teaching were held constant. Results revealed an interaction between instructor sex and both of the other factors. Behaviours indicative of friendliness toward students elevated students' ratings of instruction for female instructors, but not for male instructors. Subjects rated the male professors as more effective than female professors. Students expect female instructors to excel in both stereotypically masculine (competence) and feminine (warmth) domains. The results suggest that if female instructors want to obtain high student ratings, they must be not only highly competent with regard to factors directly related to teaching, but also careful to act in accordance with traditional sex role expectations. Male and female instructors will earn equal ratings for equal professional work only if the women also display stereotypically feminine behaviour.

Kulik, James A (1985). Using Student Ratings in Evaluating Teaching Assistants. Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 31-April 4, 1985).

The use of a computer-based system of individualized course evaluations (Instructor-Designed Questionnaires [IDQ's]) to collect student reactions to classes at the University of Michigan is discussed. Initial ratings using IDQ's were far below university norms, but ratings improved after one semester of teaching. IDQ's and similar course evaluation systems were found to provide useful information for teachers wishing to improve their teaching, but the proportion of teachers who will use this type of system voluntarily is probably small. IDQ's provide useful data to help faculty improve their performance, to help administrators, and for students choosing classes.

Marsh, H.M. (1984). "Students' Evaluation of University Teaching: Dimensionality, Reliability, Validity, Potential Biases, and Utility." Journal of Educational Psychology, 76, 707-754.

This article provides an overview of findings and research designs used to study students' evaluations of teaching effectiveness and examines implications and directions for future research. Based on this review, class-average student ratings are (a) multidimensional, (b) reliable and stable, (c) primarily a function of the instructor who teaches a course rather than the course that is taught, (d) relatively valid against a variety of indicators of effective teaching, (e) relatively unaffected by a variety of variables hypothesized as potential biases, and (f) seen to be useful by faculty as feedback about their teaching.
A broad institutional-based carefully planned program of students' evaluations of teaching effectiveness is likely to lead to the improvement of teaching. There is, however, no empirical demonstration of improvement of teaching effectiveness resulting from students' evaluations.

Marsh, Herbert W.; Overall, J.U. (1979). Validity of Students' Evaluations of Teaching: A Comparison with Instructor Self-Evaluations by Teaching Assistants, Undergraduate Faculty and Graduate Faculty. Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979.

College instructors evaluated the quality of their own teaching and were evaluated by their students in 331 different courses. Student evaluations of teaching correlated with instructor self-evaluations in courses taught by teaching assistants, in undergraduate courses taught by faculty, and in graduate level courses, demonstrating their validity at all levels of university teaching. Both student and instructor ratings were reliable, and separate factor analyses indicated that the same nine evaluation factors influenced both sets of ratings. While correlations between student and instructor ratings on the same factors were high, correlations between their ratings on different factors were low. This finding supports the distinctiveness of the different factors, as well as the use of multifactor evaluation instruments developed with the use of factor analytic techniques. These findings establish the validity and accuracy of student evaluation at all levels of university teaching, and suggest the possible usefulness of instructor self-evaluations.

McLaughlin, T.F.; and Others (1979). Instructor Feedback to Students: Effects on Subsequent Course Ratings. Paper presented at the Annual Convention of the American Psychological Association (87th, New York NY, September 1-5, 1979).

The effects of feedback letters sent to students thanking them for their course ratings were examined in two experiments. The results of both studies suggest that a simple thank you letter from the instructor can improve low ratings.

Menges, Robert J. "Evaluation in the Service of Faculty." [Chapter 4 in Seldin, Peter (1984). Changing Practices in Faculty Evaluation. San Francisco: Jossey-Bass Publishers.]

In his commentary in Seldin's book, Menges suggests that student ratings of teaching performance cannot be ignored as students have a unique opportunity to observe teaching. Further research shows that student ratings possess satisfactory reliability and validity (Overall and Marsh, 1982). Although student evaluations are not substitutes for direct measurement of learning, they do relate positively with examination performance (Cohen, 1981). Objections to student participation in evaluation deserve attention. The objections can be countered through research, judicious selection of evaluation questionnaire items, and intelligent use of results.

Murray, Harry G. (1980). Evaluating University Teaching: A Review of Research. A study commissioned and published by the Ontario Confederation of University Faculty Associations. Toronto.

Research on the reliability of student ratings of college teaching indicates that these ratings are consistent across questionnaire items, raters, time periods within a given course, and successive years for the same course; but are not consistent for different types of courses taught in the same or different years. Rating forms for student evaluation of teaching are intended to measure classroom instruction - only one measure of many components of university teaching. With respect to this limited purpose, student ratings are the best (most valid) measure available of the quality of classroom instruction. Student ratings correlate positively with ratings of independent observers and with objective measures of student learning. In most cases, teachers receiving favourable ratings from students do in fact promote higher levels of student learning than teachers receiving less favourable ratings. Student ratings are sufficiently reliable and valid to justify their use for purposes of faculty self-improvement. They are far from perfectly valid in absolute terms, and thus must be interpreted at all times with considerable caution.

Feedback from student evaluation rating forms is not a particularly effective stimulus for the improvement of college teaching. However, student feedback which is supplemented by expert consultation, motivational support and specific suggestions for improvement is more likely to lead to measureable gains in teaching performance. The author speculates that a reward system for teaching excellence might bring about improvement.

Nagel, Thomas S.; Paixao, Lyra (1971). A Study of Student Ratings of Elementary Education Instructors at San Diego State College. Paper presented at the Annual Meeting of the California Educational Research Association, San Diego, California, April, 1971.

This study investigated the differences among 65 instructors' performance according to their student ratings. It was concluded that (1) student ratings of instruction seem to be correlated with instructors' academic rank, (2) the variable tenure appears to increase differences between academic ranks, (3) the variable non-tenure appears to weaken differences between academic ranks, (4) the variable tenure, as compared with academic rank, may contribute more to discrimination between instructors, as rated by students.

Perry, R.P., Abrami, P.C., Leventhal, L., & Check, J. (1979). "Instructor Reputation: An Expectancy Relationship Involving Student Ratings and Achievement." Journal of Educational Psychology, 71, 776-787.

Instructor reputation, defined as student expectations of an instructor's teaching ability, was investigated to determine its influence on student ratings and achievement. Two teaching behaviours, instructor expressiveness and lecture content, were combined with reputation in a 2 x 2 x 2 factorial design to assess interaction effects. Results indicate that reputation interacted with expressiveness but not content. Students rated the positive, high-expressive instructor more favourably than the negative, high-expressive instructor. The results suggest a complex relationship in which reputation interacts with some teaching behaviours and affects some student outcomes. Results indicate that regardless of lecture quality, positive reputations produced higher ratings than did negative reputations. Ratings were affected by instructor reputation, but achievement was not. The consistency of the results strongly suggests that the instructor reputation effect is generalizable across a number of settings and student-ratings measures.

Scott, O. (1981). "Students' Perceptions of Locus of Control in College Classrooms and Their Global Appraisals of College Instruction." Perceptual and Motor Skills, 52, 659-664.

In this study students' perception of locus of control were conceived of as their perceptions of their roles in making decisions in the classroom. The objective was to determine if student perceptions of locus of control affect students ratings of instruction. Students' agreement or disagreement with decisions may influence their global appraisals of instruction. Instructors concerned about students' global appraisals may be well advised to pay special heed to procedures for making classroom decisions, particularly those pertaining to sequence of content and to selection of instructional materials and to provide information which can help students accept the decisions as reasonable.

Seldin, Peter (1980). Successful Faculty Evaluation Programs: A Practical Guide to Improve Faculty Performance and Promotion/Tenure Decisions. Crugers, NY: Coventry Press.

The test of reliability is to yield the same result no matter how often or where the test is performed. Almost every study regarding the reliability of student ratings has reported a high level of both correlation in time and internal stability. Student ratings once obtained tend to correlate highly with those obtained later on. Many researchers have found that questions on student rating forms interrelate well with other questions designed to measure the same instructor characteristic. "Typical measures of split-half reliability consistently show that the instrument employed in student ratings have a high reliability. They have been checked and rechecked in different studies." (p. 40)

The validity of an instrument depends on whether it measures what it is supposed to measure. The central question about validity is whether the instructors who effectively achieve educational goals are also rated highly by students. Several researchers report a moderate correlation between student ratings of instructors and their ratings by faculty colleagues. Positive correlations between student examination results and student instructor ratings have been reported.

The usefulness of student ratings in the improvement of teaching is debateable. Student ratings will produce a positive change in teaching effectiveness when discussed with the teacher by a sympathetic, knowledgeable colleague who can offer appropriate advice on means of improving teaching. Hard evidence is lacking to prove that student evaluations automatically improve classroom teaching.

Seldin, Peter (1984). Changing Practices in Faculty Evaluation. San Francisco: Jossey-Bass Publishers.

Although Seldin's research indicated that almost 70% of colleges used rating forms of one kind of another in 1983, the persistent minimal research by the institutions on the validity of the forms used to judge their faculty was noted.

Student ratings are an important element in evaluating teaching performance. The test of reliability for the data collected means that the same result is yielded no matter how often or where the test is performed. Virtually every study measuring the reliability of student ratings has reported a high level of stability and consistency. Validity means that the instrument measures what it is supposed to measure. The key to the validity of student ratings is whether the instructor rated effective by students are generally those whose students achieve most. Although study results are not uniform, it is concluded that student ratings are highly valid as indices of achievement of attitudinal and motivational goals of education and are reasonably valid as indices of achievement of cognitive goals. Age, sex, student level, and personality have scant effect on student ratings. The single most important variable affecting satisfaction is student expectation. Size of class, subject matter, and whether the class is required or elective may affect ratings to a small degree. Instructor characteristics such a sex, professional rank, and grading standards have minimal effect on student ratings.

A questionnaire whose results are to be used for formative purposes should include diagnostic questions that ask for perceptions or evaluations of specific teaching behaviours. Several open-end questions should be included to allow students to respond in their own words.

Evaluation will result in better teaching performance if (1) the appraisal is new to the teacher, (2) the teacher is motivated to improve, and (3) the teacher knows how to bring about the improvement. It is the discussion of the results that is likely to trigger changes.

Sheehan, Daniel S. (1976). "The Localization, Diagnostic, and Monitoring Functions of Student Ratings in a Model for Improving University Teaching." Instructional Science. 5, 77-92.

In this article Sheehan questions the validity of results of students' evaluation of teaching effectiveness. A diversity of student questionnaires are used by colleges and universities to provide data on faculty teaching performance. Yet the purposes for collecting this data are frequently unclear, and at times superficial. Rarely is student rating data used as a tool to improve faculty teaching. A more relevant approach incorporates a variety of types of student ratings into a model for improving university teaching. One type of student rating data (generated by the Teaching Analysis by Students (TABS) questionnaire) is used to identify broad instruction problem areas. Another type (generated by the SUB-TABS questionnaire) pinpoints probable causes and solutions for the instructional problems. Instructional improvement procedures are designed on the basis of this data. A third type of student rating data (generated by the Monitoring TABS) evaluates the instructional improvement procedures and indicates when modifications are needed. In addition to these three types of student ratings, and the generation of appropriate questionnaires, this paper presents an overview of the teaching improvement model and discusses its effectiveness.

Stevens, J.J., and Aleamoni, L.M. (1985). "The Use of Evaluative Feedback for Instructional Improvement: A Longitudinal Perspective." Instructional Science. 13, 285-304.

Both the available experimental evidence and recent reviews are unclear regarding the effectiveness of student ratings feedback for instructional improvement. Provision of feedback has not consistently produced improvement. Feedback accompanied by consultation has resulted in more consistent instructor improvement. However, the literature on interventions for instructional improvements has failed to address the longitudinal nature of improvement and development. The present study examined follow-up data over a ten-year period on instructors who had participated in a quasi-experimental study of the effectiveness of consultation and feedback conducted in 1971 and 1972. The results obtained indicate the instructors who had originally received feedback with consultation maintained higher student ratings and used instructional resources more during follow-up than instructors who originally received feedback only.

Stritter, Frank T. (1983). "Faculty Evaluation and Development." Professions Education Researcher Notes. October, 3-14

The author defines the term "teacher" and addresses concerns relating to the evaluation and development of faculty performance. When discussing the use of students' evaluation of effective teaching, Stritter suggests that this is the most typical approach to assessing an instructor's teaching performance. He stipulates that if improvement is to occur, it is essential that the instructor receive the results of evaluation while the course is in progress. The results should be accompanied by some type of systematic technical assistance in interpreting the results and in suggesting areas for improvement. The process assumes that teachers can improve their teaching when confronted with evaluation. It also assumes that teachers can improve on the basis of the feedback they receive from a systematic survey of their students.

Research on student ratings shows that they can be reliable, valid, and useful, but that they have limited generalizability. Instruments on which to obtain ratings can be developed empirically for specific faculty groups using relevant criterion variables and appropriate statistical methods, or they can be developed using a more generalizable format (Purdue University cafeteria system).

BIBLIOGRAPHY

Aleamoni, Lawrence M. (1974). Typical Faculty Concerns About Student Evaluation of Instruction. Prepared for presentation at the Symposium on Methods of Improving University Teaching at the Technion, Israel Institute o Technology, (Haifa, Israel, March 24, 1974).

Aleamoni, Lawrence M. (1972) Illinois Course Evaluation Questionnaire (CEQ) Results Interpretation Manual Form 66 and Form 32, Urbana: Office of Instructional Resources, Illinois University.

Arreola, Raoul A. (1984). "Evaluation of Faculty Performance: Key Issues." [Chapter 4 of Seldin, Peter (1984). Changing Practices in Faculty Evaluation. San Francisco: Jossey-Bass Publishers.]

Aubrecht, Judith D. "Better Faculty Evaluation Systems." [Chapter 4 of Seldin, Peter (1984). Changing Practices in Faculty Evaluation. San Francisco: Jossey-Bass Publishers.]

Bausel, R.B. & Magoon, J. (1972). "Expected Grade in a Course and Student Ratings of the Course and the Instructor." Educational and Psychological Measurement, 32, 1013-1023.

Cashin, William E.; Perrin, Bruce M. (1978). IDEA Technical Reports - Description of IDEA Standard Form Data Base. Manhattan: Center for Faculty Evaluation and Development in Higher Education.

Centra, John A. (1972). Self-Ratings of College Teachers: A Comparison with Student Ratings. Princeton, NJ: Educational Testing Service.

Chickering Arthur W. "Faculty Evaluation: Problems and Solutions." [Seldin, Peter (1984). Changing Practices in Faculty Evaluation. San Francisco: Jossey-Bass Publishers.]

Costin, Frank; Greenough, William T.; and Menges, Robert T. (1971). Student Ratings of College Teaching: Reliability, Validity, and Usefulness. Review of Educational Research. 41, 511-535.

Cross, K. Patricia, and Angelo, Thomas A. (1988). Classroom Assessment Techniques: A Handbook for Faculty. The University of Michigan: The National Center for Research to Improve Postsecondary Teaching and Learning.

Eble, Kenneth E. (1971). The Recognition and Evaluation of Teaching. Project to Improve College Teaching and The Committee on Evaluation. Salt Lake City, Utah.

Glasman, Naftaly S.; and Gmelch, Walter H. (1976). "Purposes of Evaluation of University Instructors: Definitions, Delineations, and Dimensions." Canadian Journal of Higher Education, 6, 37-55.

Gronlund, N. (1985). Measurement and Evaluation in Teaching. (5th Ed.), Collier Macmillan.

Holmes, D.S. (1972). "Effects of Grades and Disconfirmed Grade Expectancies on Students' Evaluations of Faculty Performance." Journal of Educational Psychology, 63, 130-133.

Jacobs, Lucy Cheser (1987). "University Faculty and Students' Opinions of Student Ratings." Indiana Studies in Higher Education, No. 55. Bloomington: Indiana University

Kennedy, W. Robert (1975). "Grades Expected and Grades Received - Their Relationship to Students' Evaluations of Faculty Performance." Journal of Educational Psychology, 67, 109-115.

Kierstead, Diane; D'Agostino, Patti; and Dill, Heidi (1988). "Sex Role Stereotyping of College Professors: Bias in Students' Ratings of Instructors." Journal of Educational Psychology. 880, 342-344.

Kulik, James A (1985). Using Student Ratings in Evaluating Teaching Assistants. Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 31-April 4, 1985).

Marsh, H.M. (1984). "Students' Evaluation of University Teaching: Dimensionality, Reliability, Validity, Potential Biases, and Utility." Journal of Educational Psychology, 76, 707-754.

Marsh, Herbert W.; Overall, J.U. (1979). Validity of Students' Evaluations of Teaching: A Comparison with Instructor Self-Evaluations by Teaching Assistants, Undergraduate Faculty and Graduate Faculty. Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979.

McLaughlin, T.F.; and Others (1979). Instructor Feedback to Students: Effects on Subsequent Course Ratings. Paper presented at the Annual Convention of the American Psychological Association (87th, New York NY, September 1-5, 1979).

Menges, Robert J. "Evaluation in the Service of Faculty." [Chapter 4 in Seldin, Peter (1984). Changing Practices in Faculty Evaluation. San Francisco: Jossey-Bass Publishers.]

Murray, Harry G. (1980). Evaluating University Teaching: A Review of Research. A study commissioned and published by the Ontario Confederation of University Faculty Associations. Toronto.

Nagel, Thomas S.; Paixao, Lyra (1971). A Study of Student Ratings of Elementary Education Instructors at San Diego State College. Paper presented at the Annual Meeting of the California Educational Research Association, San Diego, California, April, 1971.

Perry, R.P., Abrami, P.C., Leventhal, L., & Check, J. (1979). "Instructor Reputation: An Expectancy Relationship Involving Student Ratings and Achievement." Journal of Educational Psychology, 71, 776-787.

Scott, O. (1981). "Students' Perceptions of Locus of Control in College Classrooms and Their Global Appraisals of College Instruction." Perceptual and Motor Skills, 52, 659-664.

Seldin, Peter (1980). Successful Faculty Evaluation Programs: A Practical Guide to Improve Faculty Performance and Promotion/Tenure Decisions. Crugers, NY: Coventry Press.

Seldin, Peter (1984). Changing Practices in Faculty Evaluation. San Francisco: Jossey-Bass Publishers.

Sheehan, Daniel S. (1976). "The Localization, Diagnostic, and Monitoring Functions of Student Ratings in a Model for Improving University Teaching." Instructional Science. 5, 77-92.

Stevens, J.J., and Aleamoni, L.M. (1985). "The Use of Evaluative Feedback for Instructional Improvement: A Longitudinal Perspective." Instructional Science. 13, 285-304.

Stritter, Frank T. (1983). "Faculty Evaluation and Development." Professions Education Researcher Notes. October, 3-14

ERIC Clearinghouse for Junior Colleges
OCT 12 1990